

The Use of a Retail Store Database for Brand Choice Analysis

C. Berné, M. Cortiñas, M. Elorz and JM Múgica

Abstract

Retailers who wish to make decisions for a single store and have access to the scanner data of all purchases and to the scanner data of customer card-holders may worry about erroneous inferences when using one of the two databases and when using the same models to estimate the effects of their main marketing variables. The questions are: “do the transactions reflected by the customer cards necessarily reflect the usual purchase behavior of all customers?”; and “do the same customer response models apply equally to regular customers and for the rest of individuals shopping at the store?”. To answer these questions, a brand choice multinomial logit model choice is applied for the product category chosen (ground coffee). The major findings are that regular price elasticity of all brands in the customer segment is twice what was estimated when studying the total purchases. The effects on brand and type of coffee are greater in the customer card-holders segment than in the total purchases.

Keywords

Brand choice, regular customers, occasional shoppers, price elasticity, promotions

Introduction

Store managers may make decisions regarding product categories with the help of a choice model, and may find it very useful for pricing, promotion planning, brand selection and category assortment (brands, types, and sizes). There is extensive literature about those choice models which best reflect the effects of brands on customer preferences, price elasticity and promotion effects when using scanner data (e.g. Guadagni & Little, 1983; Gupta, 1988). In the last few years some of these models have helped to avoid problems with parameter bias due to unobserved effects (Abramson et al. 2000), to differences in the choice sets (Siddarth, Bucklin, and Morrison 1995), and also to give segmentation clues to retailers (Mazumdar & Papatla, 2000). So, from the manufacturers' perspective, the intensive research carried out in the area of marketing has provided some useful guidelines and, more interestingly, some empirical generalizations which have been incorporated to the body of their managerial knowledge (Bucklin & Gupta, 1999). But, from the retailers' perspective, are those results and empirical generalizations equally interesting and useful?

We will try to show that the databases used in these studies have some differences with the database of the individual store which may be a source of biases in the estimation of parameters in the brand choice models and that the specification of these models should be discussed before applying it to the individual store database. First, concerning the database, there has been some concern about differences in the estimated parameters of the choice model between household scanner data and store data (Gupta et al. 1996). In their study, Gupta et al. (1996), first show that “...store sales data are the sum, or aggregation, of all purchases of households (not only the households in the panel) for each store and week. Thus, the household panel brand choice data are part of the store sales data”. Based on this fact, they state that the parameters estimated for the store sales are closer to the true effects than the parameters obtained by the model estimation in the household data. With this statement, they conclude that the brand choice behavior of household panelists is statistically significantly different from the behavior represented by the store universe of all scanner-based purchases, and suggest that purchase selection (not household

selection) is best for obtaining results from household data that approximate the store data results. In summary, they take the sum of all purchases across all stores as the universe and the model estimations made on this database should be the reference to evaluate the results obtained with other samples or aggregation procedures. This might be, again, a sensible recommendation when inferences are to be made for managerial applications of manufacturers.

The individual store database is an aggregation of purchases which incorporates two sources of deviation from the true parameters which would be estimated with the store sales data. First, there is the issue of heterogeneity that has been already analyzed in different studies (Hoch et al. 1995; Montgomery, 1997) suggesting the incorporation of demographics and competition variables to control the fact that location imposes differences in the clientele and in the competitive environment. Secondly, there is another source of deviation which has not been analyzed: the coexistence, in any individual store, of regular customers (those who shop at the store with such a frequency that allows them to become familiar with the marketing and merchandising practices at the store) and other customers that can be described as occasional shoppers with little knowledge about the store selling conditions and, also, with little incentives to learn about it.

Household scanner data captures effectively the varied shopping activities of households along a period of time; households shopping for the same product category or the same brand at different retailers and different stores. More specifically, for groceries, there is common knowledge supported by research and published evidence about the shopping pattern of households. When there are available alternatives, households visit more than one grocery store regularly. Most of the households will then alternate shopping at their regular grocery store with visits to other stores of a different format and/or of a different chain.

Consequently, the transactions registered in the database of any particular supermarket capture only part of the shopping activity of its regular customers, jointly with the buying of occasional shoppers of very different profiles. This fact is often disregarded and household scanner data are converted into store data and are later used for the analysis whose results are inferred not only to manufacturers but also for retail stores (Pauwels, Hanssens, and Siddarth, 2002). Given this

substantial difference in the databases, there should be some caution in extending the inferences of the research to retail management. The joint presence of regular customers and occasional buyers generate a sales response to marketing decisions (price, promotions, assortment, displays, merchandising...) which might be substantially different if only one of the two segments is considered.

Looking at the problem from a managerial angle, retail managers who wish to make decisions for a single store and have access to the scanner data of all purchases and to the scanner data of customer card-holders, may worry about erroneous inferences when using one of the two databases. Do the transactions shown by the customer cards necessarily reflect the buying behavior of the average customer? If not, what should the implications for decision-making in the category be? The incentives linked to the use of most customer cards, support the assumption that a large part of the transactions in which the customer card is not used, is made by individuals with little shopping activity at the store for whom the use of the card would not render significant benefits. So, these individuals do not visit the store with the frequency and motivation required to become familiar with the selling conditions at the store.

Hereafter, we will name this set of transactions, those in which the customer card is not involved, as “occasional shopping”. We will show that this kind of shoppers’ heterogeneity at any store should be taken into account to avoid misleading recommendations. The first objective of the research is to provide a basis to answer these questions; by analyzing the differences that might arise when a brand choice model is estimated using total purchases and that part which corresponds to customer cards. Going one step further, if the retailer finds it logical to rely on customer card data, should decision-making be guided towards the average customer or should those customer segments with different economic weights be targeted?

The Database

For our study, we have selected coffee. This is a product category considered as storable that has been previously analyzed in empirical research (Narasimhan, Neslin, and Sen, 1996; Bell, Chiang, and Padmanabhan, 1999). The database includes all the purchases of the ground coffee category

in one large supermarket¹ over a period of one year. The coffee category chosen has a high penetration rate in households, and as such, it is expected that a very high percentage of the store customers are included in the sample.

Each purchase is one observation with the following variables: the item sold (brand, and type), the price paid, and the use or not of the customer card. Whether the purchase of one item was made on promotion or not, was inferred by the dip of price differences between two consecutive purchases of the item. The information about the use of the customer card, allows us to identify the individual cardholders. The customer card is a loyalty card that is used as a means to accumulate points, which are later redeemed in the promotions and saving plans implemented by the retailer. For the one year period studied, there were registered a total of 27,969 purchases. The number of customers who bought ground coffee with the customer card was 1,996, and the number of units purchased with the customer card was 22,771 (81.23% of total purchases). The distribution of these sales among the different brands and types, and the means of the regular price, and the promotional discounts are summarized in Table 1.

The Model

The effects of price promotions on brand choice have been intensively analyzed over the past two decades (Gupta 1988; Chintagunta 1993; Bell, Chiang, and Padmanabhan, 1999; Van Heerde, Leeflang, and Wittink, 2000) so that it has become a research topic with very solid foundations. As such, it is an adequate ground to test the hypothesized differences between the response of regular customers and occasional shoppers.

The specification of our model of brand choice tries to capture the variables that influence the choice of one of the alternatives in the product category. We include four market variables: brand, coffee type, regular price, and promotional discount. Given that all brands and types are sold in the same package format (250g), size has not been included. The brands considered were all those that were on the shelf for the 54 weeks represent 98.3% of total sales of the category. Only two brands were excluded (one brand with registered sales only in 5 weeks accounting for 1.4% of total sales, and another with only 3 weeks on the shelf accounting for 0.3% of sales). Different

types of coffee are included at this step of the research since it is not clear to what extent they are close substitutes, such that type-switching can be directly induced by changes in the other variables (brand, regular price and promotional discount).

The inclusion of prices in the model follows the conceptualization of Gupta (1988). The shelf price paid by shoppers in every transaction was transformed into regular price which is a value set up to reflect the price of the product in the absence of any discounts or promotions (Hill & Cartwright 1994). The main objective of such transformation is to separate the effect of regular price from the effect of promotional discount. Accordingly, the measurement of promotions was made by transforming the promotional price into the promotional discount as a percentage of the previous regular price. Concerning other frequently used variables, display and feature, they were not included since in our context they are redundant with the price promotion. Whenever a promotional price discount takes place, the promotion is featured and displayed in the store so that these variables are already included in the promotional discount. Another variable frequently included in the brand choice models, is loyalty (Guadagni & Little, 1983) which could not be operationalized given that the database includes not only regular customers with an available purchase history but also occasional shoppers for whom the only available information is the actual transaction.

$$P_{ij} = \frac{\exp[X_{it}'\beta]}{\sum_{j=1}^B \exp[X_{jt}'\beta]}$$

Where P_{it} is the probability of choice of brand i during week t , X_{it} is the set of marketing variables (brand, coffee type, regular price, and promotional discounts), and β is the vector of effects of marketing variables. A descriptive view of the model variables for the total sales of the category and for the two segments, regular customers and occasional shoppers, is provided in Table 1.

The Hypotheses

The model was estimated separately for the total database, for the segment of customer card transactions, and for the segment of occasional shoppers. The segments were identified by means of the use of the customer card in the transaction. Whenever an item was bought with the use of

a customer card, it was assigned to the regular customer segment. The basis for this assumption is that only customers with a regular shopping activity in the store can find benefits with the use of a customer card. The segment of occasional shoppers was made up by all other transactions in which the customer card was not present. It is more risky to describe this segment as made up only by occasional shoppers since it is likely that some of the transactions included correspond to regular customers misusing the card. However, it is expected that most of these transactions correspond to consumers that, for a variety of reasons, do not have a regular shopping activity at the store.

Regular customers have a critical difference with occasional shoppers. They have knowledge about the brand and type assortment, the shelf prices at which the items are usually sold, and the promotions' dynamics. Regular customers are expected to be familiar with the different brands usually sold at the store so that very likely they have developed a more clear preference hierarchy.

Hypothesis 1

Brands and types will have a higher impact in the choice of regular customers than in the choice of occasional shoppers. The visits to the category shelf, the exposition to features, displays, and other media used by the retailer to communicate the prices are the bases to build up a solid frame of reference for price evaluation purposes. The measurement of the price variable, the regular price, is in fact an Internal Reference Price (IRP) because it implies that consumers have a memory of past prices of a brand. Regular customers are expected to have a moderate accurate knowledge about the prices charged by the retailer so that they have a certain ability to evaluate not only the price fluctuations of their most preferred brands but, also, the relative differences between their prices. On the other hand, occasional shoppers lack this frame of reference. They might be more familiar with prices at other stores or not even that. A similar consideration has to be made for the measurement of the promotional discount. In our model, it has been measured as a percentage discount relative to the regular price so that it presumes that all consumers make their choices based on information about past prices. In other conceptualizations (Gupta, 1988) it

is not a percentage but the absolute value of the discount, which also implies knowledge of the past price.

Hypothesis 2

The regular price parameter and the specific brand elasticities in a brand choice model estimated for the total transactions of the category are biased by the presence of occasional shoppers. Regular price is not a valid reference for brand choice in the segment of occasional shoppers.

Hypothesis 3

If the measurement of the promotion variable implies a discount evaluation relative to past prices, the promotional discount parameter in a brand choice model estimated for the total transactions of the category is biased by the presence of occasional shoppers. Discounts relative to past prices or to regular prices are not a valid reference for brand choice in the segment of occasional shoppers.

To test the hypotheses, the model was estimated for the total database, for the sample of customers, and for the sample of occasional buyers. As differences in the estimation show up, the results will be discussed to assess the validity of the initial model for the total database and for the two segments.

The Results

The estimation of the model for the total database and the two segments is provided in Table 2. Prior to the test of the hypotheses, the estimation reflects large differences between the two segments and it is clear that the model fits much better and is more explanatory, measured by the probability ratio index, in the segment of regular customer transactions, than for the segment of occasional shoppers. Also, all the parameters show better significance ratios. This confirms the initial assumption that the random component in the choice behavior of occasional shoppers is larger. Also, the estimation shows that the transactions of occasional shoppers, although they are a small proportion of the total purchases in the category, have a strong effect on the estimates of the parameters of the total purchases database. So it is indicating that the participation of

occasional shoppers in the category sales is biasing the parameters of the variables in the choice model.

More specifically, the first two hypotheses are confirmed: brand preferences (heterogeneity) are much more differentiated in the customer segment, and the regular price is not relevant for the occasional shoppers' segment. First, regarding brand preferences, the range of the parameter from the least to the most preferred brand is 2.378, which is 80% larger than the 1.324 range observed in the occasional shoppers segment. This range indicates that the differentiation perceived and valued by regular customers in their choices in the store is higher. Also, there is a relevant change in the market structure. Although the two leader brands, Bonka and Marcilla, hold the two top positions in both segments, the brands which could be considered as followers looking at their position in the total database are Soley, Brasilia and Saimaza (with very similar market shares close to the 8%), change their position in the occasional shoppers segment. Saimaza drops to the last position and is well differentiated from the rest of the brands while the least preferred brand in the customer segment, Bahia, moves up two positions. The preference for types of coffee is also affected by the inclusion of the occasional shopper segment. The natural not-blended type is the least preferred in the occasional shopper segment. As in the case of brands, store customers take in to consideration the types of coffee much more than the occasional shoppers. The range width is 1.391 in the customer segment which is two and a half times longer than in the occasional shoppers' segment (0.397).

The hypothesis concerning the regular price parameter is partially confirmed. First, the comparison between the regular price parameter for the total database with the parameter estimated for the customer segment indicates that the presence of occasional shoppers is biasing downwards the estimation. If no attention is paid to the effect of occasional shoppers and only regular customers are considered, the parameter that would be estimated is twice the actual one (0.12 in the regular customer segment vs. 0.06 in the total database). Consequently, the regular price elasticity of all brands would be also overestimated if the occasional shoppers were not considered (Table 3).

Although the parameter of regular price is significant in the two segments, the effect on brand choice is almost irrelevant in the case of occasional shoppers (0.004). The fact that it has a positive sign, against what would be expected, might be explained because many choices in the segment are made randomly, and/or taken without a valid price reference framework.

Hypothesis 3 is not confirmed. The estimated parameters of the promotional discount are of a very similar magnitude for the two segments. One plausible explanation is that the absence of past prices' information for occasional shoppers is not absolute. The promotional display may include the absolute or relative magnitude of the discount with respect to the previous price. Also, the simple fact that a brand is featured and displayed as being in promotion is signaling a discount to any shopper.

Conclusions

The objectives of this research involved the answers to the following questions: “do the transactions reflected by the customer cards necessarily reflect the actual purchase behavior of all customers?”, and “do the same customer response models apply equally to regular customers and for the rest of individuals shopping at the store?”.

In relation to the first question, it has been found that the brand choice behavior of occasional shoppers is very different from that of regular customers and that it is a potential source of bias in the estimation of parameters in the brand choice model. More specifically, it has been found that brand preferences (heterogeneity) are much more differentiated in the regular customer segment, and that the regular price is not relevant for the occasional shoppers' segment. This last finding explains that the regular price elasticity of all brands, in the regular customer segment (active card-holders), is twice what was estimated when studying the total purchases. One of the basic managerial implications for retailers is that if they design a pricing and promotions policy linked to the customer card, they should be aware that the response will be to diverge from that one predicted by the estimations which use the database of total transactions. This divergence will become larger with increases in the percentage of transactions made by occasional shoppers.

In relation to the second question, we suggest that the brand choice model to be applied to occasional shoppers should be modified to account for the peculiarities of their shopping activity in the store. There is no reason to believe that the individuals that do not use the customer card at the store have different shopping behavior or demand characteristics than the regular customers. Their general behavior, for the total of their grocery shopping activity, should not be explained in different terms. It is only in that small part of their shopping activity, the transaction out of their main store, in which their behavior might not be determined in the same way as for regular customers.

The modeling of these transactions should first acknowledge that the random component is larger and cannot be explained at the same extent as for regular customer's transactions by the observed variables. In fact, the modeling efforts should not try to diminish the random component, as it is inherently larger.

Also, some facts should be considered when taking decisions about the measurement of the variables. First, occasional shoppers do not know what past prices were at the store, and are not sensitive to regular price variations. There is no reason to believe that they react differently to any other regular shopper to a promotional discount display. So, variable measurement in the brand choice model to be applied to occasional shoppers should be modified to account for the peculiarities of their shopping activity in the store. The basic peculiarity that is assumed to play an important role in their choice behavior is that they are choosing brands separately from their price reference frame and from their habitual brand choice set. This fact lessens the consistency of their decisions from the perspective of the retailer. Thus, the two modifications suggested are to substitute the regular price variable by the shelf price, and to substitute the promotional discount with a dummy variable simply indicating if the brand purchased was on promotion or not. It is expected that occasional shoppers make some use of the price information displayed on the shelves (shelf prices) and that might be attracted rather by the presence of a price promotion than for the amount of discount.

It is necessary to know the structure of the database of the company, discerning the customer cardholders segment from the rest. The derived decisions, mainly about pricing, will then be made in a real context.

References

- Abramson, C., Andrews, R. L., Currim, I. S. & Jones M. (2000) Parameter bias from unobserved effects in the multinomial logit model of consumer choice, *Journal of Marketing Research*, 37(November), pp. 410–426.
- Bell, D. R., Chiang, J. & Padmanabhan, V. (1999) The decomposition of promotional response: an empirical generalization, *Marketing Science*, 18(4), pp. 504–26
- Bucklin, R. E. & Gupta, S. (1999) Commercial use of UPC scanner data: industry and academic perspectives, *Marketing Science*, 18(3): 247–273.
- Chintagunta, P. K. (1993) Investigating purchase incidence, brand choice and purchase quantity decisions of households, *Marketing Science*, 12(2), pp. 184–208.
- Guadagni, P. M. & Little, J. D. (1983) A logit model of brand choice calibrated on scanner data, *Marketing Science*, 2(3), pp. 203–238.
- Gupta, S. (1988) Impact of sales promotions on when, what, and how much to buy, *Journal of Marketing Research*, 25(November), pp. 342–355.
- Gupta S., Chintagunta P., Kaul A. & Wittink D. R. (1996) Do household scanner data provide representative inferences from brand choices: a comparison with store data, *Journal of Marketing Research*, 33(4), pp. 383–398.
- Hill, R. C. & Cartwright, P. (1994) The statistical properties of the equity estimator, *Journal of Business and Economic Statistics*, 12(April), pp. 141–153.
- Hoch, S. J., Kim, B., Montgomery, A. M. & Rossi, P. E. (1995) Determinants of store-level price elasticity, *Journal of Marketing Research*, 32(February), 17–29.
- Mazumdar, T. & Papatla, P. (2000) An investigation of reference price segments, *Journal of Marketing Research*, 37(2), pp. 246–258.
- Montgomery, A. (1997) Creating micro-marketing pricing strategies using supermarket scanner data, *Marketing Science*, 16(4), pp. 315–337.

- Narasimhan, C., Neslin, S. & Sen, S. (1996) Promotional elasticities and category characteristics, *Journal of Marketing*, 60(April), pp. 17–30.
- Pauwels, K., Hanssens, D. M. & Siddarth, S. (2002) The long-term effects of price promotions on category incidence, brand choice, and purchase quantity, *Journal of Marketing Research*, 39(November), pp. 421–439.
- Siddarth, S., Bucklin, R. E. & Morrison, D. G. (1995) Making the cut: modeling and analyzing choice set restriction in scanner panel data, *Journal of Marketing Research*, 32(August), pp. 255–266.
- Van Heerde, H. J., Leeflang, P. S. & Wittink, D. R. (2000) The estimation of pre- and post-promotion dips with store-level scanner data, *Journal of Marketing Research*, 37(August), pp. 383–395.

Tables

Table 1. Descriptive values of the variables

		Total		Regular Customers		Occasional Shoppers	
Variables		Sales (units)	Share (%)	Sales (units)	Share (%)	Sales (units)	Share (%)
Brand	Bonka	9,595	34.31	7,820	34.34	1,775	34.15
	Marcilla	8,532	30.51	7,008	30.78	1,524	29.32
	Soley	2,501	8.79	2,191	9.62	268	5.16
	Brasilia	2,459	8.94	2,152	9.45	349	6.71
	Saimaza	2,411	8.62	1,780	7.82	631	12.14
	M156	1,394	4.98	1,141	5.01	253	4.87
	Bahia	1,077	3.85	679	2.98	398	7.66
Type	Blend	17,476	62.48	14,639	64.29	2,837	54.58
	Natural	8,196	29.30	6,874	30.19	1,322	25.43
	Special	2,297	8.21	1,258	5.52	1,039	19.99
Total		27,969	100	22,771	81.42	5,198	18.58
Regular Price mean (RP)			1.33 e		1.32 e		1.35 e
Shelf price mean (SP)			1.23 e		1.23 e		1.26 e
Promotional discount mean (PD)			12.86%		12.96%		12.44%
Units purchased on promotion			54.74 %		54.87%		54.19 %

Table 2. Estimation of the brand choice model for the total database and for the two segments

		Parameters (t values)		
Variables		Total purchases	Regular customers	Occasional shoppers
Brands	Bonka	1,926 (52,812)	2,378 (52,814)	0,677 (9,989)
	Marcilla	1,803 (48,572)	2,316 (49,685)	0,433 (6,660)
	Soley	1,652 (42,693)	2,125 (45,513)	0,232 (2,797)
	Brasilia	0,974 (23,002)	1,345 (26,524)	0,090 (1,072)
	Saimaza	0,767 (17,506)	1,306 (24,430)	-0,647 (-7,469)
	M156	0,226 (5,599)	0,342 (6,911)	-0,245 (-3,367)
	Bahia	Base	Base	Base
Type	Blend	1,056 (40,656)	1,391 (42,892)	0,106 (2,155)
	Natural	0,622 (22,757)	0,965 (28,693)	-0,290 (-5,420)
	Special	Base	Base	Base
Regular price (RP)		-0,006 (-10,767)	-0,012 (-17,677)	0,004 (3,936)
Promotional Discount (PD)		9,629 (97,665)	9,990 (89,740)	8.819 (39,373)
LL		-62.929	-49.908	-12.303
r ₂₀		0,147	0,169	0,103
r ₂₀		0,135	0,157	0,090

Table 3. Comparison of the brand specific elasticities

		Total Database		Regular Customer	
		Reg. price	Prom. discount	Reg. price	Prom. discount
Brands	Bonka	−0.859	0.192	−1.717	0.219
	Marcilla	−0.908	0.203	−1.810	0.231
	Brasilia	−1.192	0.266	−2.363	0.302
	Soley	−1.190	0.266	−2.368	0.302
	Saimaza	−1.194	0.267	−2.410	0.308
	M156	−1.242	0.278	−2.484	0.317
	Bahia	−1.257	0.281	−2.537	0.324
Types	Blend	−0.490	0.110	−0.934	0.119
	Natural	−0.924	0.207	−1.825	0.233
	Special	−1.200	0.268	−2.470	0.315